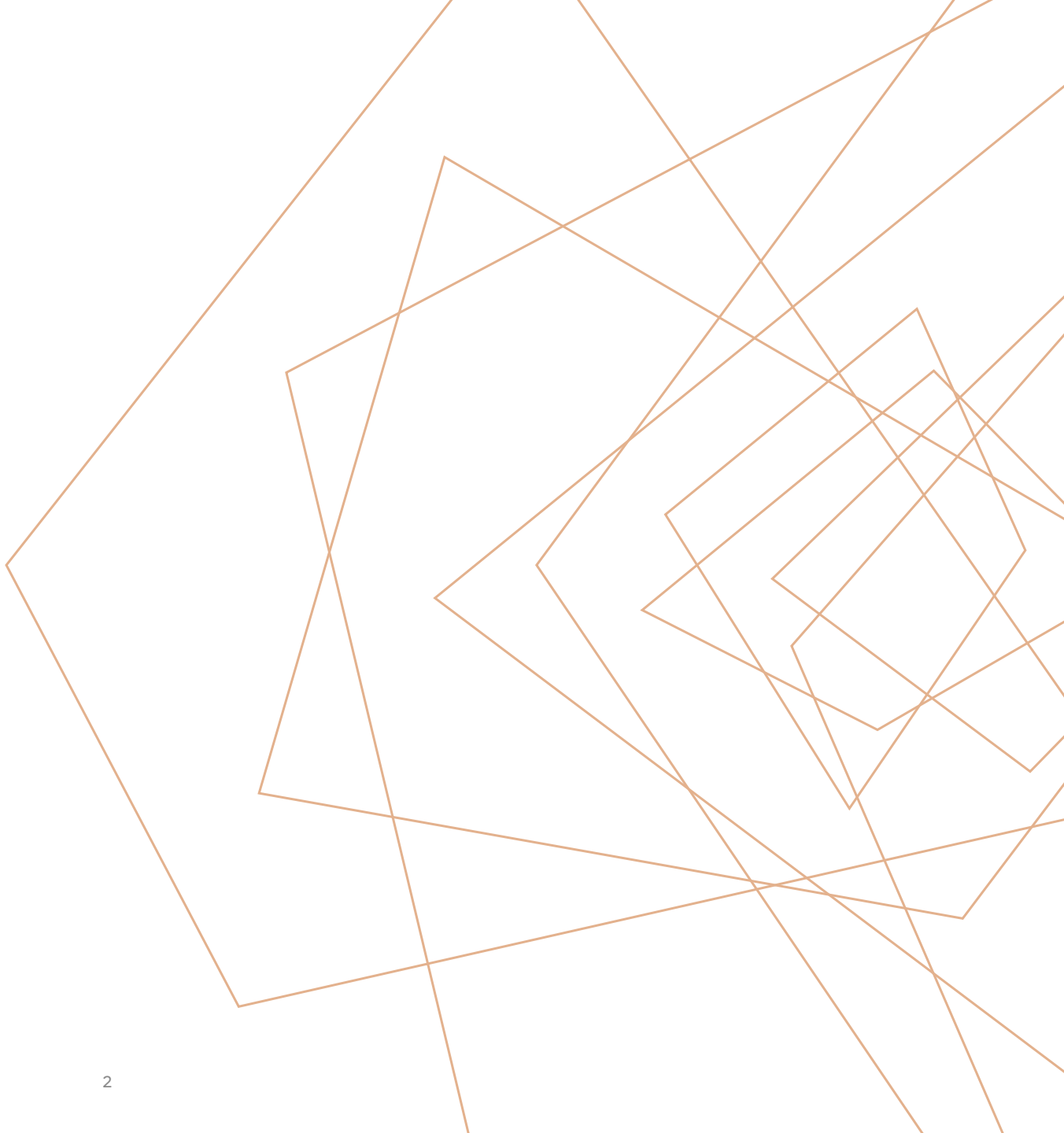
LEAD SCORING CASE STUDY

Onkar Suryawanshi

Manish Mishra

INTRODUCTION

An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. The company markets its courses on several websites and search engines like Google. Once these people land on the website, they might browse the courses or fill up a form for the course or watch some videos. When these people fill up a form providing their email address or phone number, they are classified to be a lead.



PROBLEM STATEMENT

X Education sells online courses to industry professionals

X Education gets a lot of leads through website, its lead conversion rate is very poor. For example, if, say, they acquire 100 leads in a day, only about 30 of them are converted.

To make this process more efficient, the company wishes to identify the most potential leads, also known as ‘Hot Leads’.

If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

BUSINESS OBJECTIVE

X education wants to know most promising leads.

For that they want to build a Model which identifies the hot leads.

Deployment of the model for the future use.

If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone.

SOLUTION

DATA CLEANING & DATA MANIPULATION

- Importing the data and to check the duplicate
  - Handling of Null & Missing Values
  - Dropping of unnecessary columns
  - Handling the outliers in the data

EDA

- Perform Univariate data analysis
  - Perform Bivariate data analysis

DUMMY VARIABLES & ENCODING OF DATA

LOGISTIC REGRESSION FOR MODEL MAKING & PREDICTION

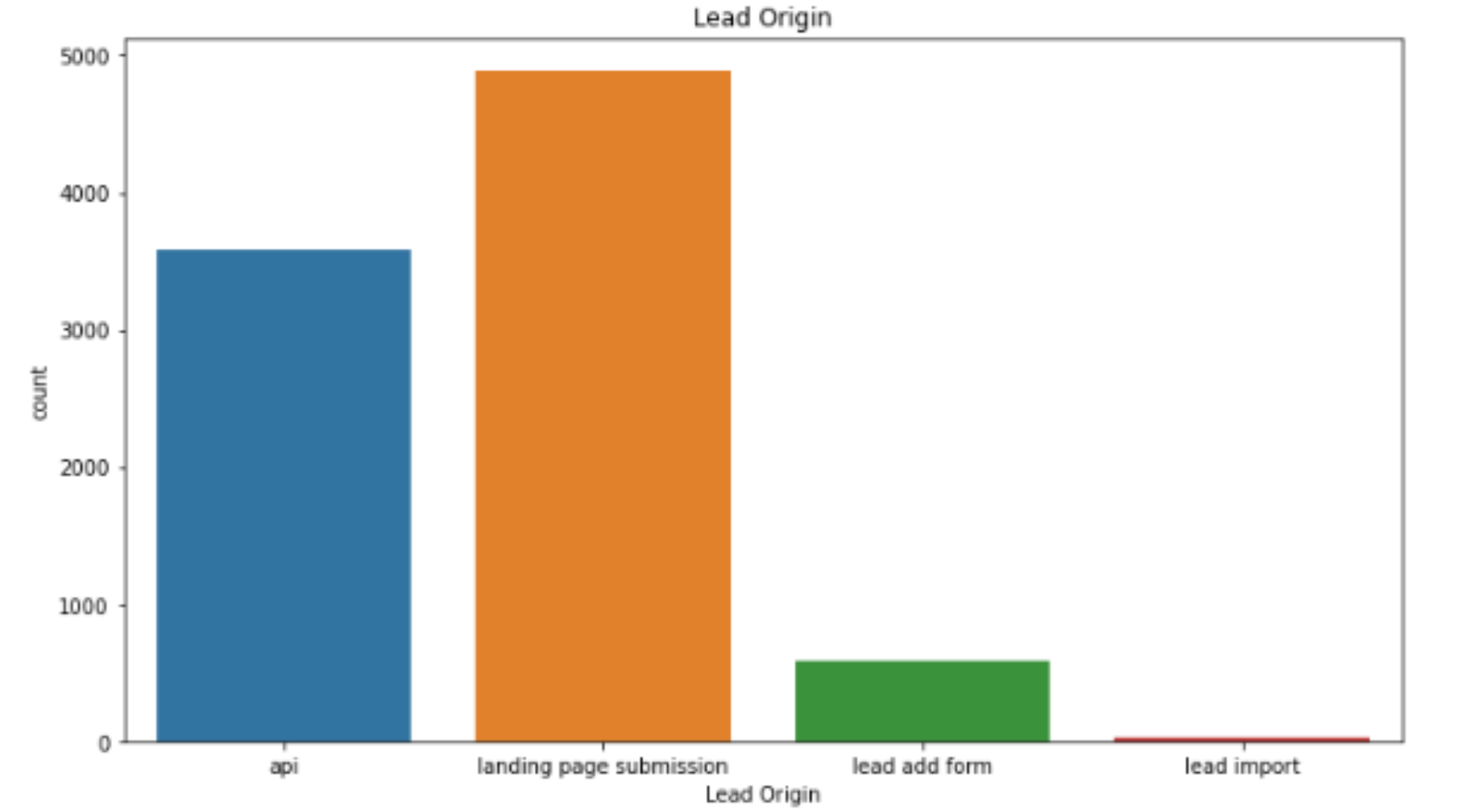
MODEL VALIDATION

MODEL PRESEANTATION

RECOMMENDATIONS

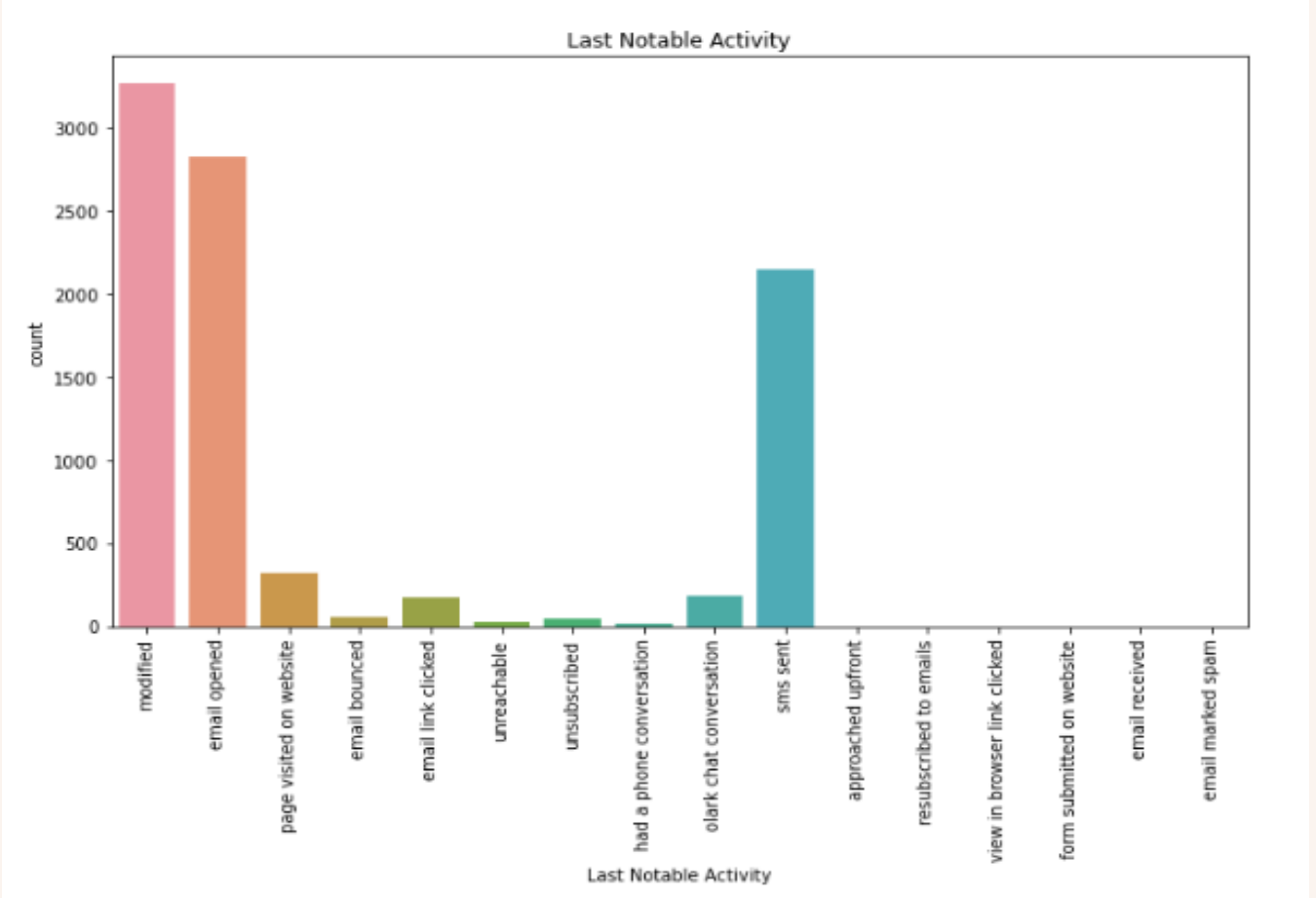
EXPLORATORY DATA ANALYSIS

Forecasting for success



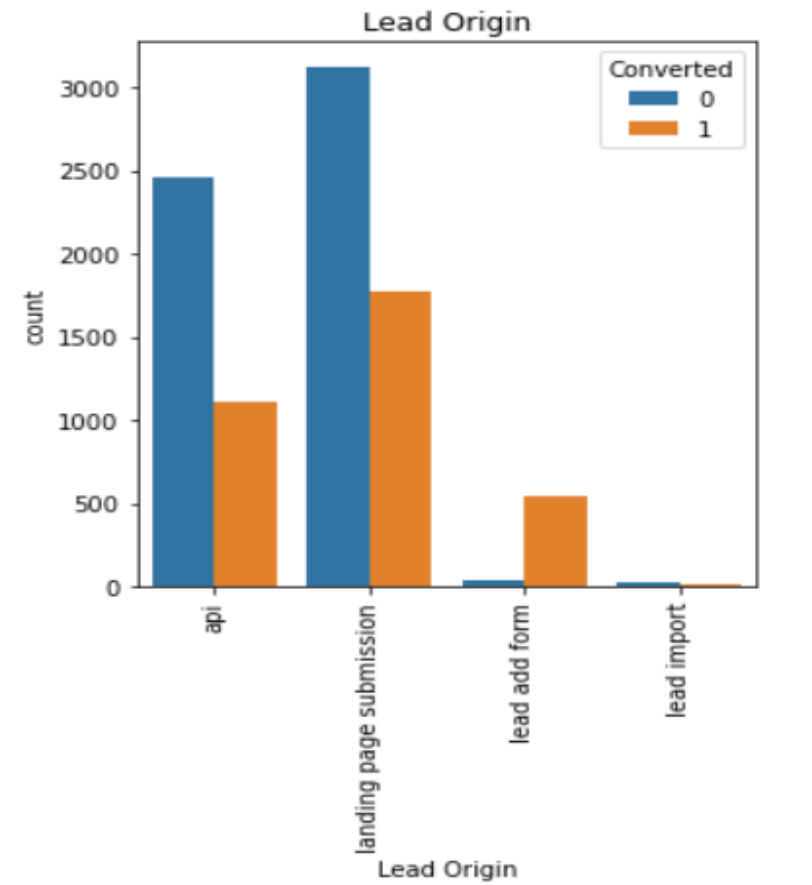
EXPLORATORY DATA ANALYSIS

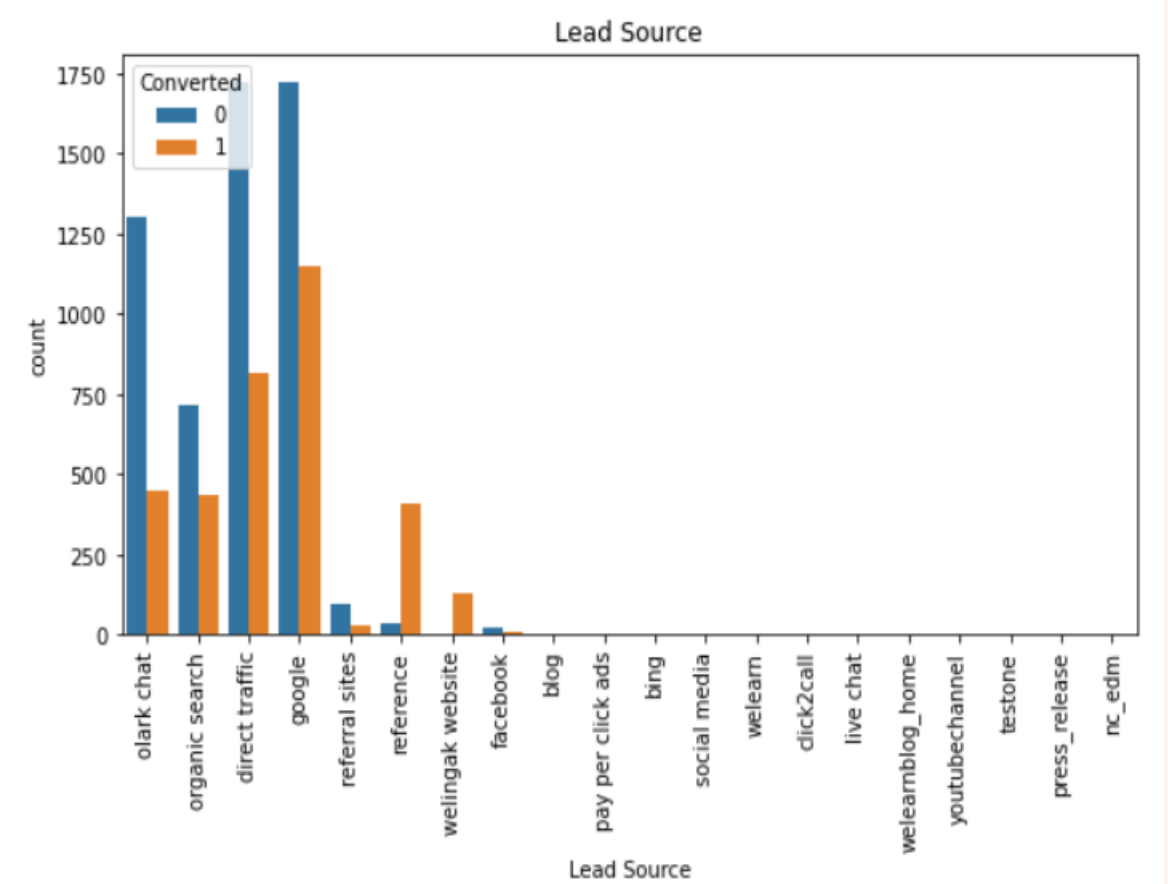
Observing the behavioral pattern of the by observing the last notable Activity



EXPLORATORY DATA ANALYSIS

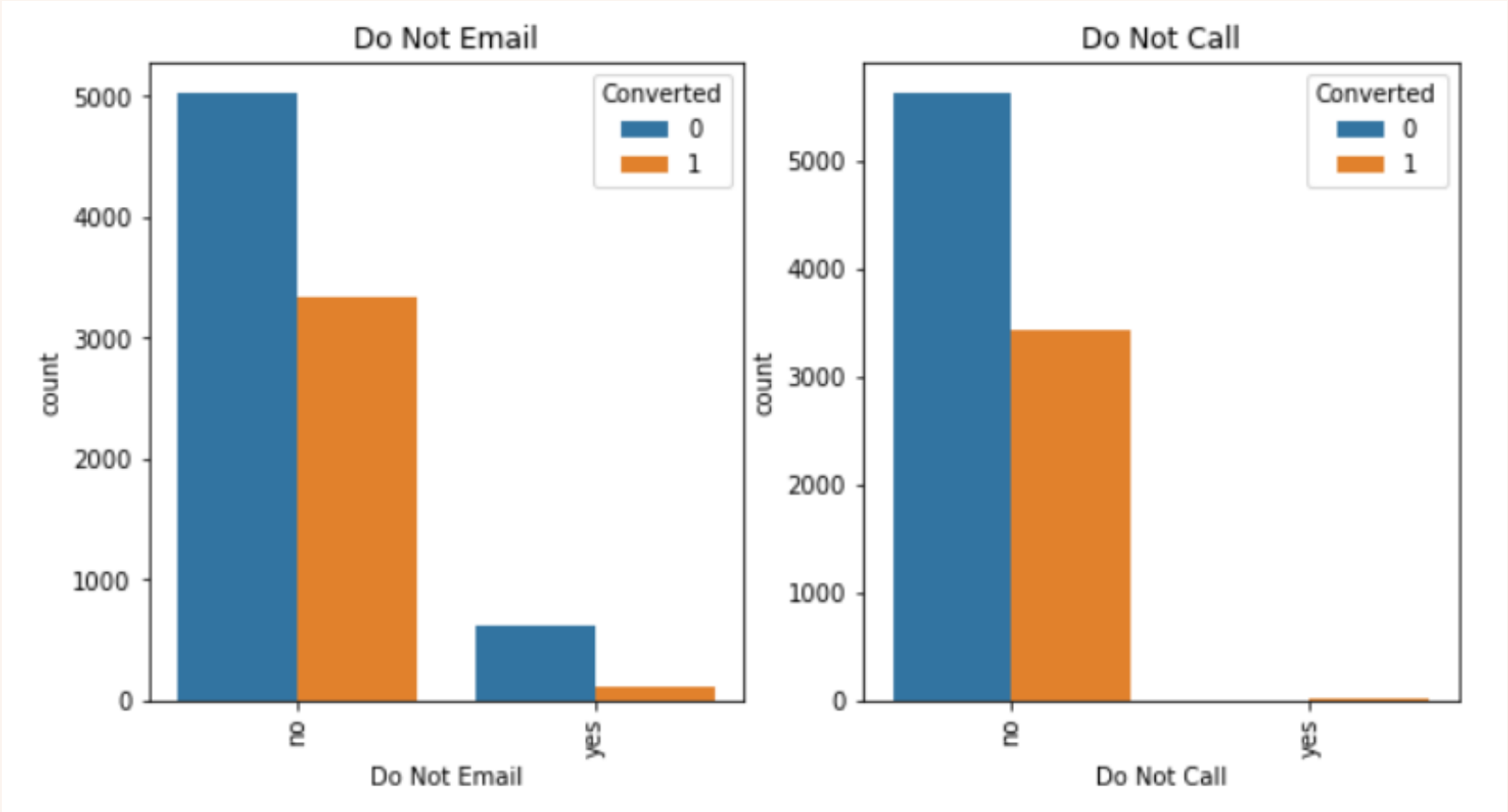
CATEGORICAL VARIABLE VS TARGET VARIABLE





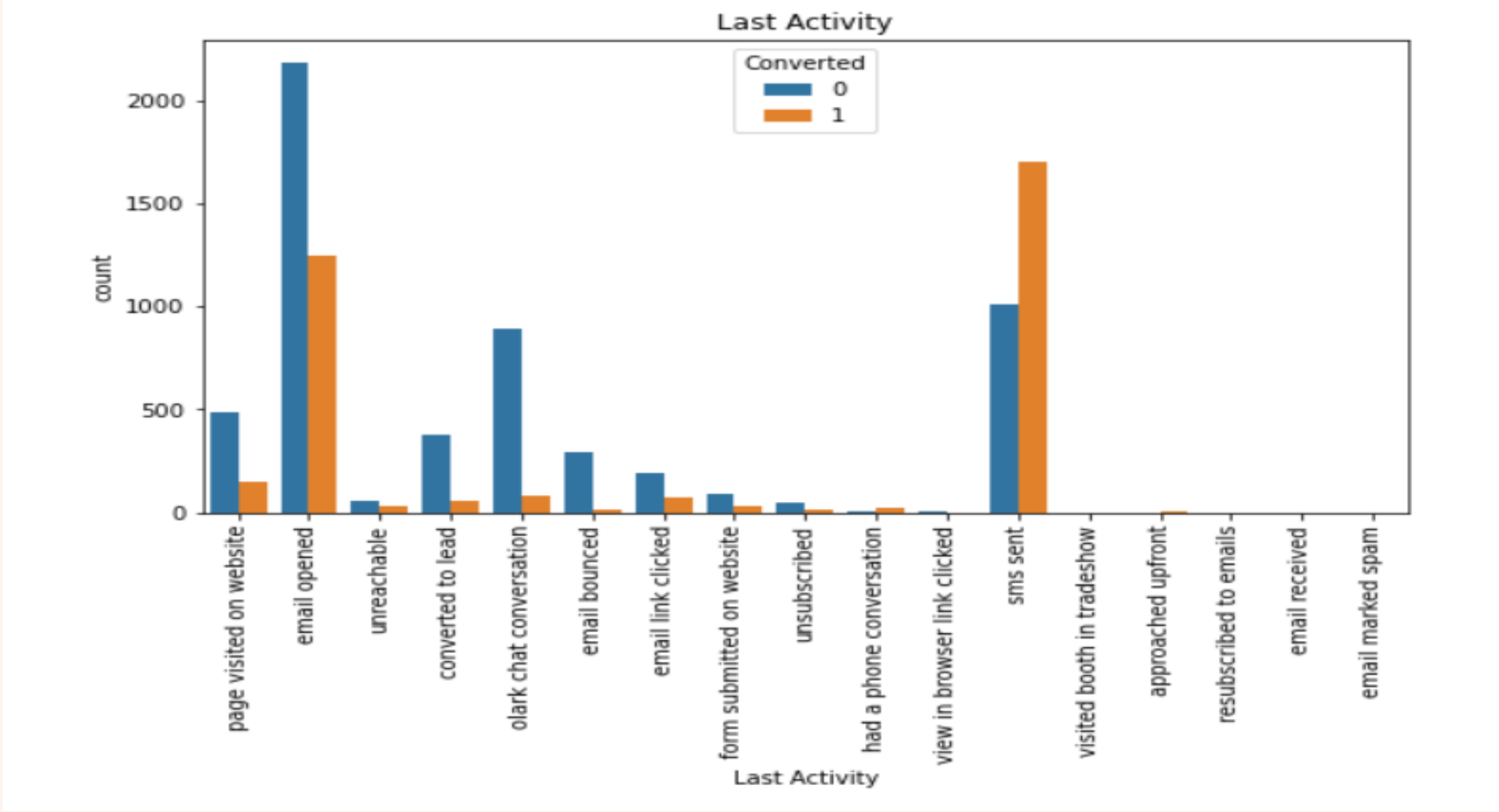
EXPLORATORY DATA ANALYSIS

CATEGORICAL VARIABLE VS TARGET VARIABLE



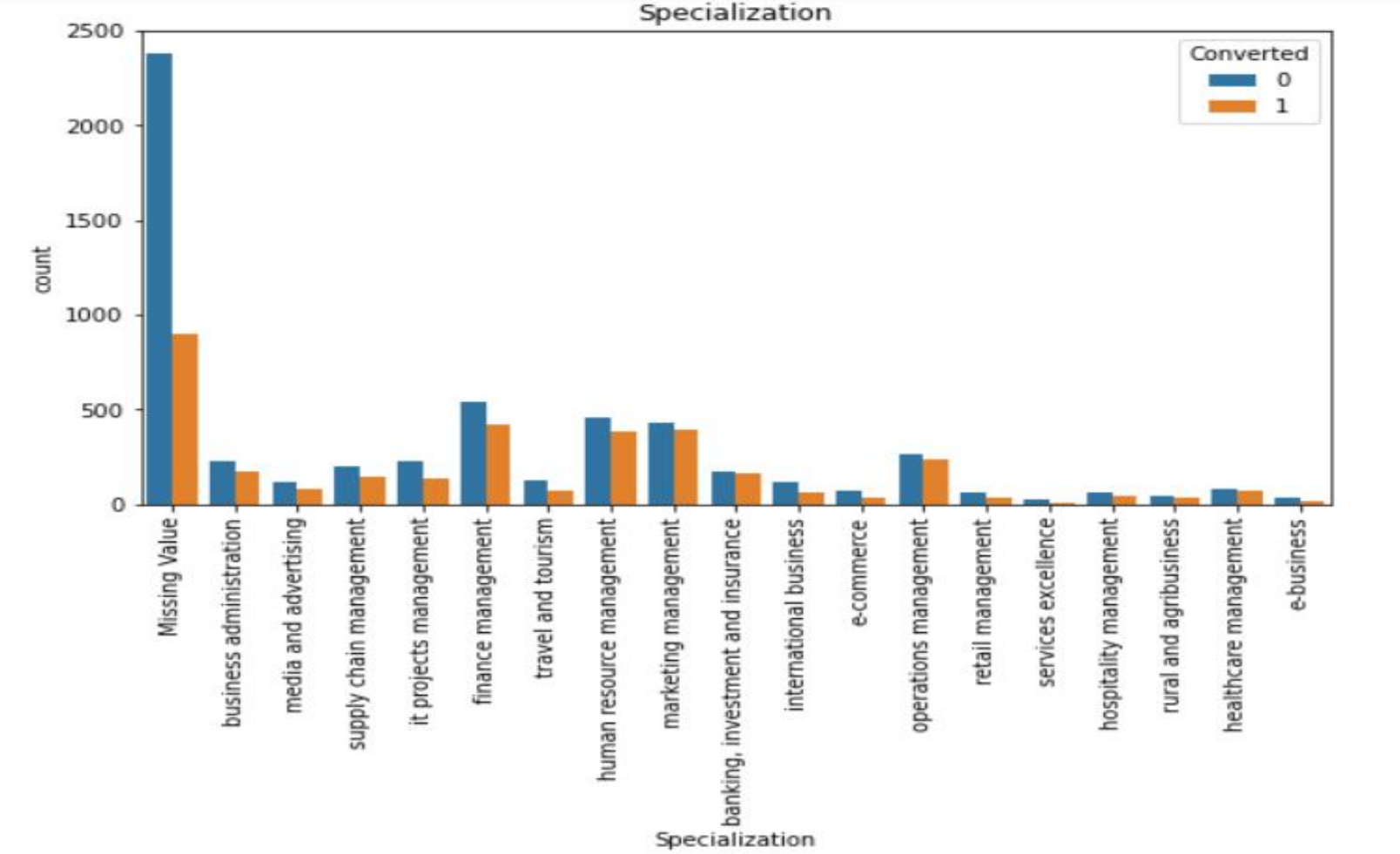
EXPLORATORY DATA ANALYSIS

CATEGORICAL VARIABLE VS TARGET VARIABLE



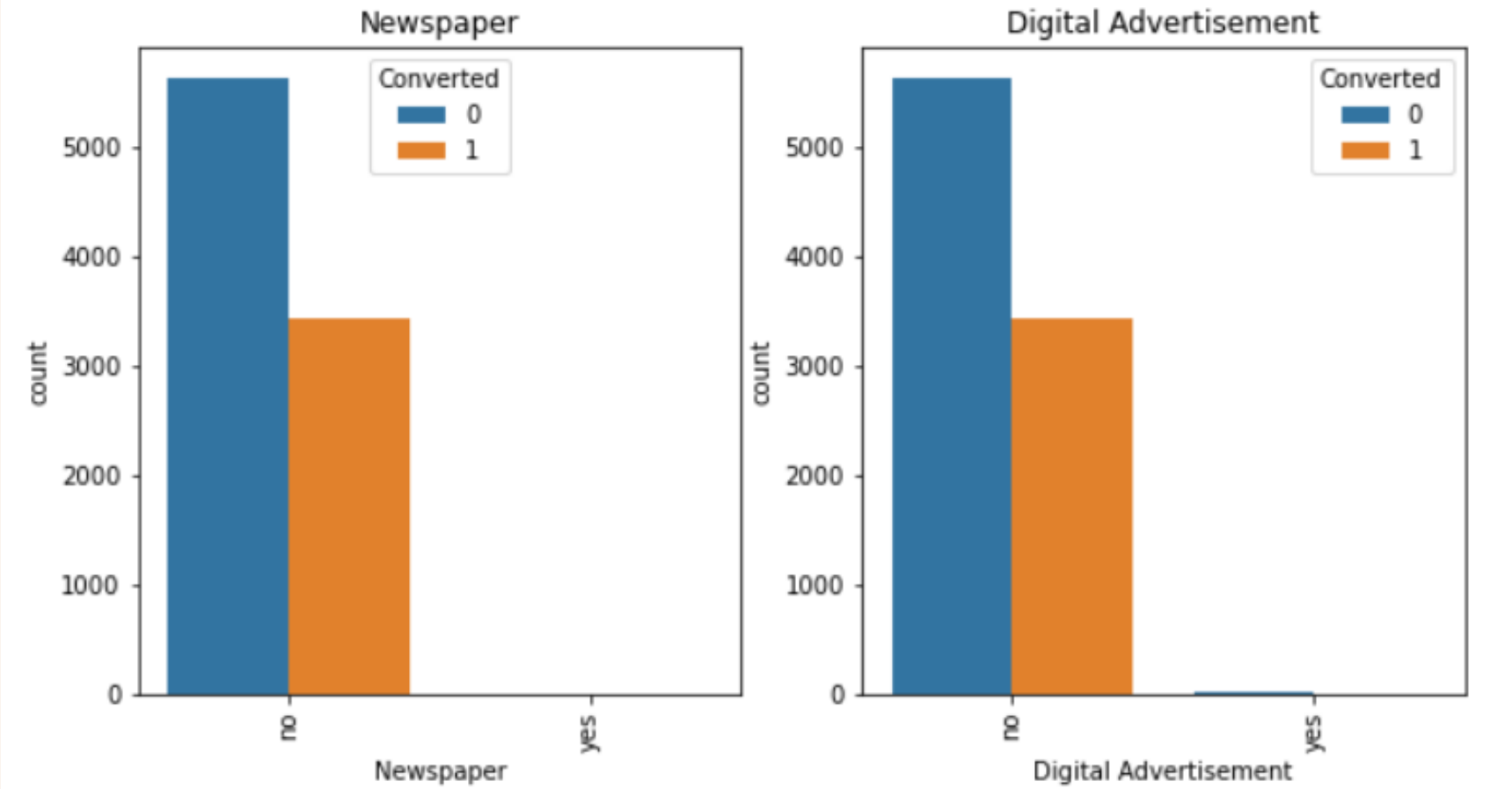
EXPLORATORY DATA ANALYSIS

CATEGORICAL VARIABLE VS TARGET VARIABLE



EXPLORATORY DATA ANALYSIS

CATEGORICAL VARIABLE VS TARGET VARIABLE



Total Rows for Analysis – 9074

Total Columns for Analysis – 100

Dummy variables are created for object type variables

Numerical variables are normalized

DATA CONVERSION

MODEL BUILDING

Split the data into training & test data sets

Basic step for regression is to perform train test split ratio for that is 70:30

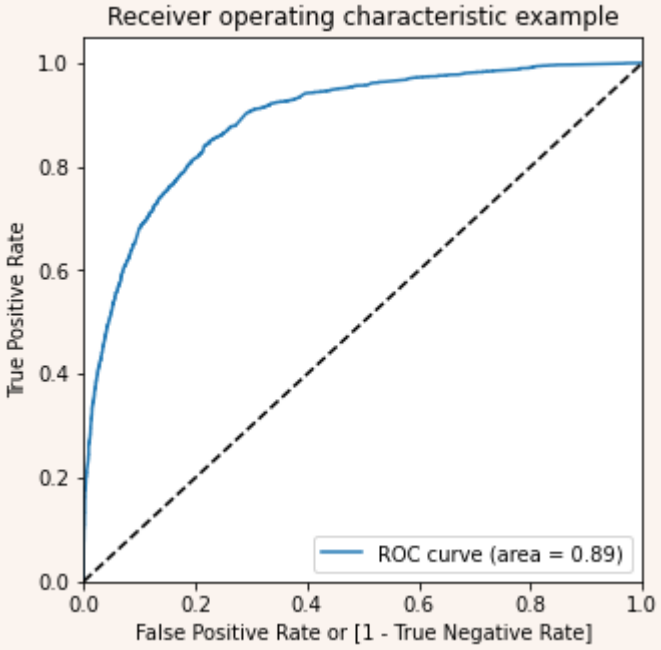
RFE feature selection for selecting most relevant in predicting the target variable.

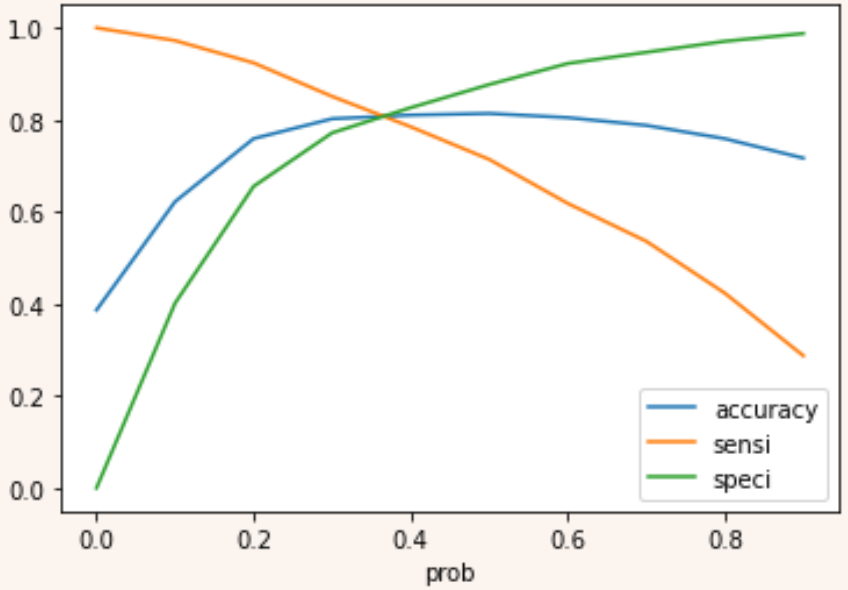
Building the model

Predictions on test data set

Overall Accuracy – 81.89%

ROC CURVE





- Need to find optimal cut off point
- ROC curve is 89%

- Need to get balanced sensitivity & specificity
- The optimal cut off is 0.5

CONCLUSION

As we can see after doing analysis on the data we came to the conclusion that which matter in serious buyers

- The total time spend on the website
- Total number of visits
- The leads source came from various platforms like Google, Website, etc.
- The leads come repeatedly on website
- The website user interface should be interactive so the leads could be engaged.

THANK YOU

Onkar Suryawanshi

Manish Mishra